Archives of Academic Emergency Medicine. 2026; 14(1): e24

REVIEW ARTICLE

Revisiting the Diagnostic Accuracy of Ottawa Ankle Rules for Ankle Fractures: A Systematic Review and Meta-Analysis

Danial Sharifi Razavi[1] Mohammadali Seiri[2], Mehrdad Farrokhi[3,4]*

1. Department of Orthopedics, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran
2. Department of Orthopedics, School of Medicine, Iran University of Medical Sciences, Tehran, Iran
3. ERIS Research Institute, Tehran, Iran
4. Department of Epidemiology, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Received: March 2026; Accepted: May 2026; Published online: 18 June 2026

Abstract: **Introduction:** The Ottawa Ankle Rules (OAR) have demonstrated conflicting results across studies, with generally high sensitivity but relatively low and variable specificity. This updated meta-analysis aims to evaluate the overall diagnostic characteristics of the OAR for early identification of ankle fractures in adult patients. **Methods:** In this systematic review and meta-analysis, a comprehensive literature search was conducted from database inception through May 2026 across three electronic databases, including Medline, Web of Science, and Scopus. The MIDAS package in STATA and Meta-Disc software were used to pool the findings of the diagnostic accuracy studies. Clinical application of the OAR was assessed using Fagan's nomogram and scattergram. **Results:** The pooled sensitivity and specificity of OAR were 0.92 (95% confidence interval (CI): 0.91–0.93) and 0.35 (95% CI: 0.34–0.36), respectively. The pooled positive likelihood ratio (PLR) and negative likelihood ratio (NLR) were 1.76 (95% CI: 1.46–2.13) and 0.13 (95% CI: 0.09–0.19), respectively. Furthermore, the pooled diagnostic odds ratio (DOR) was 16.21 (95% CI: 10.15–25.89). **Conclusion:** In conclusion, the OAR demonstrate excellent sensitivity and a very low negative likelihood ratio, confirming their value as a reliable screening tool for ruling out ankle fractures in clinical practice. With pooled sensitivities exceeding 90% and a post-test probability reduced to approximately 1% following a negative result, the OAR can safely reduce unnecessary radiographic imaging.

Keywords: Ankle Fractures; Ankle Injuries; Clinical Decision Rules; Meta-analysis

Cite this article as: Sharifi Razavi D, Seiri M, Farrokhi M. Revisiting the Diagnostic Accuracy of Ottawa Ankle Rules for Ankle Fractures: A Systematic Review and Meta-Analysis. Arch Acad Emerg Med. 2026; 14(1): e24. https://doi.org/10.22037/aaem.v14i1.3051.

1. Introduction

Ankle injuries, which demonstrate a high prevalence, represent the second most common form of lower-extremity trauma presenting to emergency departments and are among the most frequently encountered conditions managed by emergency clinicians (1). Moreover, in the United States, approximately 14.6% of the 117 million emergency department visits were reported to involve injuries primarily affecting the lower extremities (2). Following physical examination, imaging, most commonly plain ankle radiography, constitutes the primary diagnostic modality for the early detection of ankle injuries, including fractures and sprains (3). Nevertheless, the routine use of radiographic imaging for all patients presenting with ankle injuries contributes to unnecessary radiation exposure, increased burden on emergency and radiology departments, and elevated healthcare expenditures (4, 5). These concerns underscore the importance of reducing unnecessary imaging, particularly among patients with a low probability of ankle fracture (5). The Ottawa Ankle Rules (OAR), a well-established clinical decision-making tool originally introduced by Stiell et al. (6), is widely utilized by healthcare professionals to reduce the number of unnecessary ankle radiographs through the early exclusion of ankle fractures (7). These clinical criteria have been disseminated across various countries as a decision-support tool for clinicians, with the aim of minimizing exposure to ionizing radiation and reducing patient burden in emergency departments through the decreased use of unnecessary radiographic imaging (5, 6). However, the OAR have demonstrated conflicting results across studies, with generally high sensitivity but relatively low and variable specificity (5, 8, 9). This variability underscores the importance of conducting secondary research to synthesize and pool the findings of existing studies. Therefore, this updated meta-analysis aims to evaluate the overall diagnostic characteristics of the OAR for the early identification of ankle fractures in adult patients.

***Corresponding Author:** Mehrdad Farrokhi; Department of Epidemiology, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran. Email: drfarrokhi2024@gmail.com, Tel: 00989120566704, ORCID: https://orcid.org/0009-0004-6638-8017.

2. Methods

2.1. Search strategy

The protocol for this meta-analysis was registered in PROSPERO (CRD420261404225). A comprehensive literature search was conducted from database inception through May 2026 across three electronic databases, including Medline, Web of Science, and Scopus. Keywords related to the OAR, ankle injuries, and diagnostic performance, along with Medical Subject Headings (MeSH) terms for PubMed, were combined using Boolean operators (Appendix 1). To minimize the likelihood of missing eligible studies, supplementary search strategies were employed, including gray literature searches and manual screening of the reference lists and citation records of the included studies.

2.2. Eligibility criteria

We included studies that reported sufficient data to construct a 2 × 2 contingency table for evaluating the diagnostic accuracy of the OAR in detecting ankle fractures among adult patients. The analysis focused specifically on ankle fractures; however, studies reporting combined diagnostic data for both ankle and foot fractures were also included when separate ankle-specific results were unavailable. Studies conducted exclusively in pediatric populations or involving participants younger than 18 years of age were excluded. In addition, case reports, case series, conference abstracts, preprints, and other non–peer-reviewed publications were excluded from the analysis.

2.3. Data extraction and quality appraisal

Two authors independently extracted both diagnostic and baseline study data, and any discrepancies were resolved through consultation with a third author. The extracted variables included the first author's name, percentage of male participants, mean age, country of study, reference standard, sensitivity, specificity, and data required for the construction of 2 × 2 contingency tables. The methodological quality and risk of bias of the included studies were assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool.

2.4. Statistical analyses

The MIDAS package in STATA and Meta-Disc software were used to pool the findings of the diagnostic accuracy studies. Publication bias was assessed using Deeks' funnel plot asymmetry test. Outlier detection and influence analyses were conducted to identify potentially influential or outlying studies. In addition, Fagan's nomogram and scattergram analyses were employed to evaluate the clinical applicability of the OAR for the diagnosis of ankle fractures.

3. Results

3.1. Characteristics of the include studies

A systematic search of the three databases yielded a total of 98 studies, of which 21 were removed as duplicates using EndNote. Title and abstract screening of the remaining 77 articles resulted in 53 studies being considered eligible for full-text review. Following full-text evaluation, 43 studies met the eligibility criteria and were ultimately included in the meta-analysis, as illustrated in the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) flow diagram (Figure 1).

A total of 43 original studies involving 16,503 patients were included in the meta-analysis. Two studies were entered twice in the analysis, with one study providing separate data for geriatric and non-geriatric populations and the other reporting independent prospective and retrospective evaluations. The included studies were published between 1993 and 2025. The majority of the studies used radiography as the reference standard, while one study additionally utilized three-dimensional computed tomography (3D CT) alongside radiography as the gold standard. Further diagnostic and baseline characteristics of the included studies are summarized in Table 1.

3.2. Quality appraisal and publication bias

The main methodological concerns among the included studies were related to the risk of bias in the patient selection domain, followed by concerns regarding flow and timing, as well as applicability concerns associated with patient selection (Table 2). Furthermore, because many studies did not clearly report whether the assessors of the index test were blinded to the results of the reference standard, and vice versa, these studies were classified as having an unclear risk of bias.

Asymmetry test of Deeks' funnel plot revealed no publication bias ($p = 0.79$) (Figure 2).

3.3. Meta-analyses

The pooled sensitivity and specificity were 0.92 (95% confidence interval (CI): 0.91–0.93) and 0.35 (95% CI: 0.34–0.36), respectively (Figure 3). The pooled positive likelihood ratio (PLR) and negative likelihood ratio (NLR) were 1.76 (95% CI: 1.46–2.13) and 0.13 (95% CI: 0.09–0.19), respectively (Figure 4). Furthermore, the pooled diagnostic odds ratio (DOR) of the OAR was 16.21 (95% CI: 10.15–25.89) (Figure 5). The area under the curve (AUC) derived from the summary receiver operating characteristic (SROC) curve was 0.87 (Figure 5). The pooled diagnostic accuracy estimated using the metaprop package was 0.52 (95% CI: 0.45–0.59) (Figure 6). The forest plot shows that sensitivity estimates were consistently high across most studies, indicating strong ability of the test to identify ankle fractures, whereas specificity demonstrated substantial variability between studies. Compared with the observed estimates, the empirical Bayes es-

timates were more stabilized and shrunk toward the pooled mean, reflecting improved precision through hierarchical modeling (Figure 6).
The goodness-of-fit plot reveals a sigmoidal deviation from the diagonal reference line, with deviance residuals forming a flat plateau in the lower quantiles before rising sharply, suggesting suboptimal model fit and potential heterogeneity among studies (Figure 7). Similarly, the bivariate normality plot shows reasonable adherence to the reference line in the lower quantiles, but notable upper-tail divergence indicates that a subset of studies violates the bivariate normality assumption. The outlier detection plot identifies several studies flagging them as statistical outliers. Similarly, the boxplot analysis identified several studies as potential outliers among the included studies (Figure 8). Fagan's nomogram demonstrated that a positive result on the OAR increases the probability of ankle fracture from 20% to 29%, whereas a negative result reduces this probability to approximately 1% (Figure 8). The likelihood ratio scattergram placed the OAR in the left lower quadrant, suggesting a primary role in ruling out ankle fractures rather than confirming them (Figure 8). Similarly, the probability-modifying plot indicated that a negative result from the OAR is more informative, as the curve approaches the point (1, 0) and substantially reduces the pre-test probability of ankle fractures (Figure 9).
The sensitivity and specificity of the OAR for the detection of ankle fractures, based on studies that exclusively reported ankle fracture data or provided sufficient data for separate extraction of ankle fracture outcomes, were 0.91 (95% CI: 0.90–0.92) and 0.34 (95% CI: 0.33–0.35), respectively (Figure 9).
The Grading of Recommendations, Assessment, Development and Evaluation (GRADE) assessment indicated moderate certainty of evidence for both sensitivity and specificity sections (Table 3).

3.4. Meta-regression

Meta-regression was performed using the following covariates: age group (≥35 years vs. <35 years), population type (geriatric vs. non-geriatric), clinical setting (emergency department vs. other settings), geographic region (developing vs. developed countries), study design (prospective vs. retrospective), assessor type (physician vs. other or mixed assessors), and fracture type (ankle and foot fractures vs. ankle fractures only) (Figure 9).
The covariates "clinical setting" and "geographic region" were excluded from the final meta-regression analysis because the software failed to achieve model convergence, resulting in a maximum iteration error during model estimation. Meta-regression analysis demonstrated that only fracture type was significantly associated with diagnostic performance.

4. Discussion

Given inconsistencies across previous studies and the absence of an updated meta-analysis incorporating recent evidence, the present study aimed to evaluate the diagnostic performance of the OAR for the detection of ankle fractures. The pooled sensitivity and specificity of OAR were 0.92 and 0.35, respectively. The positive likelihood ratio (PLR) and negative likelihood ratio (NLR) were estimated at 1.76 and 0.13, respectively. The area under the summary receiver operating characteristic (SROC) curve was 0.87, and the overall diagnostic accuracy was 0.52. These findings suggest that, due to its high sensitivity and low NLR, OAR is more clinically useful for ruling out ankle fractures rather than confirming their presence. However, a substantial proportion of the included studies demonstrated a high risk of bias, particularly in the domains of patient selection and flow and timing, while several studies were also rated as having unclear risk of bias. Therefore, these limitations warrant cautious interpretation of the pooled estimates.
In a similar meta-analysis conducted by Gomes et al. (8), the literature search was performed up to 2020 and yielded 15 studies with a total sample size approximately half that of the present meta-analysis. They reported a pooled sensitivity of 0.91, which was comparable to the estimate observed in our analysis, whereas the pooled specificity of 0.25 was considerably lower than our estimated specificity. An important methodological difference between the two meta-analyses relates to the eligibility criteria of the included studies. In the study by Gomes et al., only studies reporting diagnostic performance specifically for ankle injuries were included, and in studies presenting combined diagnostic parameters for ankle and midfoot injuries, ankle-specific data were extracted separately. In contrast, the present meta-analysis also included studies reporting combined diagnostic parameters for ankle and midfoot injuries, as the majority of patients in those studies presented primarily with ankle trauma. Another notable difference concerns the methodological quality assessment of the included studies. Gomes et al. did not classify any study as having a high risk of bias in the patient selection domain, whereas in our assessment, 21 studies were judged to have a high risk of bias for patient selection. This issue may limit the generalizability of the findings to the broader emergency department population, as participants in several studies were selected based on restrictive exclusion criteria and may not fully reflect routine clinical practice.
Interestingly, in our meta-analysis, when the sensitivity and specificity of the OAR for the detection of ankle fractures were estimated using only studies that exclusively reported ankle fracture data or studies from which ankle-specific data could be extracted separately from combined ankle and foot fracture analyses, similar pooled estimates were observed, with a sensitivity of 0.91 and a specificity of 0.34. In this regard, a meta-analysis by Bachmann et al. (7) evaluated the diagnostic performance of the OAR for the exclusion of ankle and

midfoot fractures. They performed subgroup analyses separately for studies focusing on ankle fractures and those reporting combined foot and ankle fractures. In contrast to our findings, their results showed lower sensitivity and specificity in studies including combined foot and ankle fractures compared with studies restricted to ankle fractures.

This discrepancy may be partly explained by differences in inclusion criteria and the number of included studies, as their analysis also incorporated pediatric populations and was based on a smaller number of eligible publications.

It has been shown in another meta-analysis that the Ottawa Ankle and Foot Rules (OAFR) and the OAR demonstrate similar diagnostic performance, with sensitivity values of 0.93 versus 0.94 and specificity values of 0.45 versus 0.42, respectively (10). Although their reported sensitivity was comparable to that observed in our study, the specificity was higher than that obtained in our meta-analysis. Notably, that meta-analysis included both adult and pediatric populations, whereas the present study was limited to adult patients only. In a meta-analysis by Dowling et al. (11), the diagnostic accuracy of the OAR for detecting ankle and midfoot fractures in pediatric populations was assessed. The review included 12 studies comprising 3,130 patients identified through searches of Medline, CINAHL, Embase, and the Cochrane Library. They reported a pooled sensitivity of 98.5%, which was considerably higher than the sensitivity of 0.92 observed in the present meta-analysis. Accordingly, the higher sensitivity reported in previous meta-analyses that included both adult and pediatric populations, compared with our findings, may be partly explained by the greater diagnostic sensitivity of the OAR in pediatric patients.

5. Limitations

Our study has several limitations. First, the literature search was restricted to studies published in the English language, which may have introduced language bias. Second, we did not include or compare the diagnostic performance of the OAR between pediatric and adult populations. Third, studies exclusively evaluating midfoot fractures were not included; therefore, direct comparison of the diagnostic performance of the OAR for midfoot versus ankle fractures was not possible. Finally, most of the included studies used plain radiography as the reference standard, despite the fact that this imaging modality may not accurately detect all ankle fractures.

6. Conclusions

In conclusion, the OAR demonstrated excellent sensitivity and a very low negative likelihood ratio, confirming its value as a reliable screening tool for ruling out ankle fractures in clinical practice. With pooled sensitivities exceeding 90% and a post-test probability reduced to approximately 1% following a negative result, the OAR can safely reduce unnecessary radiographic imaging. However, the consistently low specificity and modest positive likelihood ratio indicate limited ability to confirm fractures, meaning that positive OAR findings should be interpreted cautiously and followed by radiographic evaluation rather than used as definitive evidence of fracture. Furthermore, the observed heterogeneity, evidence of outlier studies, and deviations from model assumptions suggest variability in study populations, clinical settings, and application of the OAR across studies. Despite these limitations, the findings consistently support the OAR as an effective exclusion tool, particularly in emergency and primary care settings where the primary clinical objective is to safely identify patients who do not require imaging.

7. Declarations

7.1. Acknowledgments

None.

7.2. Authors' contributions

All authors contributed to study design, data collection, and writing the draft of the study. All authors read and approved final version of manuscript.

7.3. Funding/Support

None.

7.4. Data Availability

Not applicable.

7.5. Conflict of interest

None.

7.6. Using Artificial Intelligence Chatbots

We used ChatGPT version 5.1 to check grammar and enhance the academic quality and clarity of the text, which was primarily written by the authors. The final version of the text, after incorporating ChatGPT-assisted improvements, was thoroughly reviewed and approved by the authors.

References

1. Lambers K, Ootes D, Ring D. Incidence of patients with lower extremity injuries presenting to US emergency departments by anatomic region, disease category, and age. Clin Orthop Relat Res. 2012;470(1):284-90.
2. Niska R, Bhuiya F, Xu J. National Hospital Ambulatory Medical Care Survey: 2007 emergency department summary. Natl Health Stat Report. 2010(26):1-31.
3. Matharu G, Najran P, Porter K. Soft-tissue ankle injuries. Trauma. 2010;12 (2):105 - 15.
4. Berrington de González A, Darby S. Risk of cancer from diagnostic X-rays: estimates for the UK and 14 other countries. Lancet. 2004;363(9406):345-51.
5. Beckenkamp PR, Lin CC, Macaskill P, Michaleff ZA, Maher CG, Moseley AM. Diagnostic accuracy of the Ottawa

Ankle and Midfoot Rules: a systematic review with meta-analysis. Br J Sports Med. 2017;51(6):504-10.
6. Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell I, Worthington JR. A study to develop clinical decision rules for the use of radiography in acute ankle injuries. Ann Emerg Med. 1992;21(4):384-90.
7. Bachmann LM, Kolb E, Koller MT, Steurer J, ter Riet G. Accuracy of Ottawa ankle rules to exclude fractures of the ankle and mid-foot: systematic review. Bmj. 2003;326(7386):417.
8. Gomes YE, Chau M, Banwell HA, Causby RS. Diagnostic accuracy of the Ottawa ankle rule to exclude fractures in acute ankle injuries in adults: a systematic review and meta-analysis. BMC Musculoskelet Disord. 2022;23(1):885.
9. Gomes YE, Chau M, Banwell HA, Davies J, Causby RS. Adequacy of clinical information in X-ray referrals for traumatic ankle injury with reference to the Ottawa Ankle Rules-a retrospective clinical audit. PeerJ. 2020;8:e10152.
10. Barelds I, Krijnen WP, van de Leur JP, van der Schans CP, Goddard RJ. Diagnostic Accuracy of Clinical Decision Rules to Exclude Fractures in Acute Ankle Injuries: Systematic Review and Meta-analysis. J Emerg Med. 2017;53(3):353-68.
11. Dowling S, Spooner CH, Liang Y, Dryden DM, Friesen C, Klassen TP, et al. Accuracy of Ottawa Ankle Rules to exclude fractures of the ankle and midfoot in children: a meta-analysis. Acad Emerg Med. 2009;16(4):277-87.
12. Ahmad I, Rehman S, ur Rehman I, Qayyum SF, Siraj N. Diagnostic accuracy of Ottawa rules in diagnosing ankle fractures among patients taking X ray as gold standard. Professional Med J. 2025;32(05):522-7.
13. Ak R, Kurt E, Bahadırlı S, Semih Çakır M, Bilgü AS, Kurt Ş ZE. The Comparison of Ottawa Ankle Rules and Shetty test performances in foot-ankle trauma patients visited to the emergency department. Injury. 2022;53(6):2287-91.
14. Anyanwu E, Madubueze C, Alada A, Umaru H, Dzong B, Sharif A. Validation of the Ottawa Ankle Rule in Blunt Ankle and Midfoot Injuries in a Trauma Care Centre in Nigeria. J West Afr Coll Surg. 2025;15(4):394-9.
15. Auleley GR, Kerboull L, Durieux P, Cosquer M, Courpied JP, Ravaud P. Validation of the Ottawa ankle rules in France: a study in the surgical emergency department of a teaching hospital. Ann Emerg Med. 1998;32(1):14-8.
16. Avinca Ö, Taş M. Comparison of Shetty ankle test and Ottawa ankle rules in ankle injuries. Eur Rev Med Pharmacol Sci. 2023;27(5):1852-5.
17. Beceren GN, Yolcu S, Tomruk O, Atay T, Baykal YB. Ottawa versus Bernese: which is better? Eur J Trauma Emerg Surg. 2013;39(2):147-50.
18. Broomhead A, Stuart P. Validation of the Ottawa Ankle Rules in Australia. Emerg Med (Fremantle). 2003;15(2):126-32.
19. Can U, Ruckert R, Held U, Buchmann P, Platz A, Bachmann LM. Safety and efficiency of the Ottawa Ankle Rule in a Swiss population with ankle sprains. Swiss Med Wkly. 2008;138(19-20):292-6.
20. Daş M, Temiz A, Çevik Y. Implementation of the Ottawa ankle rules by general practitioners in the emergency department of a Turkish district hospital. Ulus Travma Acil Cerrahi Derg. 2016;22(4):361-4.
21. Derksen RJ, Knijnenberg LM, Fransen G, Breederveld RS, Heymans MW, Schipper IB. Diagnostic performance of the Bernese versus Ottawa ankle rules: Results of a randomised controlled trial. Injury. 2015;46(8):1645-9.
22. Dwivedi R, Ale S. Evaluation of the accuracy of “Ottawa ankle rules” for predicting fractures in acute ankle and midfoot injuries. J Univ Coll Med Sci. 2014;2(2):2-5.
23. Fiesseler F, Szucs P, Kec R, Richman PB. Can nurses appropriately interpret the Ottawa Ankle Rule? Am J Emerg Med. 2004;22(3):145-8.
24. Glas AS, Pijnenburg BA, Lijmer JG, Bogaard K, de RM, Keeman JN, et al. Comparison of diagnostic decision rules and structured data collection in assessment of acute ankle injury. Cmaj. 2002;166(6):727-33.
25. Jia CQ, Mbori NJR, Li S, Dai Y, Xu F, Porter DE, et al. Ottawa ankle and foot rules in China: applicability in a defensive environment. Eur J Med Res. 2025;30(1):687.
26. Kelly AM, Richards D, Kerr L, Grant J, O'Donovan P, Basire K, et al. Failed validation of a clinical decision rule for the use of radiography in acute ankle injury. N Z Med J. 1994;107(982):294-5.
27. Keogh SP, Shafi A, Wijetunge DB. Comparison of Ottawa ankle rules and current local guidelines for use of radiography in acute ankle injuries. J R Coll Surg Edinb. 1998;43(5):341-3.
28. Leisey J. Prospective validation of the Ottawa Ankle Rules in a deployed military population. Mil Med. 2004;169(10):804-6.
29. Lucchesi GM, Jackson RE, Peacock WF, Cerasani C, Swor RA. Sensitivity of the Ottawa rules. Ann Emerg Med. 1995;26(1):1-5.
30. Makiev KG, Vasios IS, Keskinis A, Moustafa RM, Petkidis G, Ververidis A, et al. Shetty Test Challenges Ottawa Ankle Rules in Detecting Foot and Ankle Fractures: A Prospective Comparative Study. Bull Emerg Trauma. 2025;13(1):20-4.
31. Marinelli M, Di Giulio A, Mancini M. Validation of the Ottawa ankle rules in a second-level trauma center in Italy. J Orthop Traumatol. 2007;8(1):16-20.
32. McBride KL. Validation of the Ottawa ankle rules. Experience at a community hospital. Can Fam Physician. 1997;43:459-65.
33. Morais B, Branquinho A, Barreira M, Correia J, Machado M, Marques N, et al. Validation of the Ottawa ankle rules: Strategies for increasing specificity. Injury. 2021;52(4):1017-22.
34. Murphy J, Weiner DA, Kotler J, McCormick B, Johnson D, Wisbeck J, et al. Utility of Ottawa Ankle Rules in an Aging

Population: Evidence for Addition of an Age Criterion. J Foot Ankle Surg. 2020;59(2):286-90.
35. Özbay H, Yüksel S. Clinical usefulness of the Ottawa Ankle Rules in the overweight and obese population following an acute ankle injury: A prospective cross-sectional study. Acta Orthop Traumatol Turc. 2021;55(5):435-8.
36. Papacostas E, Malliaropoulos N, Papadopoulos A, Liouliakis C. Validation of Ottawa ankle rules protocol in Greek athletes: study in the emergency departments of a district general hospital and a sports injuries clinic. Br J Sports Med. 2001;35(6):445-7.
37. Perry S, Raby N, Grant PT. Prospective survey to verify the Ottawa ankle rules. J Accid Emerg Med. 1999;16(4):258-60.
38. Pigman EC, Klug RK, Sanford S, Jolly BT. Evaluation of the Ottawa clinical decision rules for the use of radiography in acute ankle and midfoot injuries in the emergency department: an independent site assessment. Ann Emerg Med. 1994;24(1):41-5.
39. Pijnenburg AC, Glas AS, De Roos MA, Bogaard K, Lijmer JG, Bossuyt PM, et al. Radiography in acute ankle injuries: the Ottawa Ankle Rules versus local diagnostic decision rules. Ann Emerg Med. 2002;39(6):599-604.
40. Pires R, Pereira A, Abreu ESG, Labronici P, Figueiredo L, Godoy-Santos A, et al. Ottawa ankle rules and subjective surgeon perception to evaluate radiograph necessity following foot and ankle sprain. Ann Med Health Sci Res. 2014;4(3):432-5.
41. Rosin A, Sinopoli M. Impact of the Ottawa Ankle Rules in a U.S. Army troop medical clinic in South Korea. Mil Med. 1999;164(11):793-4.
42. Salt P, Clancy M. Implementation of the Ottawa Ankle Rules by nurses working in an accident and emergency department. J Accid Emerg Med. 1997;14(6):363-5.
43. Santelli F, Magnanelli R, Pallotto R, Formica L, Polonara S, Marinelli M. Validation of the Ottawa Ankle Rules in athletes: Study in a II level Trauma Center. Med Sport. 2008;61 (3):373-80.
44. Seyhan AU, Ak R, Şimşek F, Ayvacı S, Açıkgöz O. The role of Ottawa ankle rules in geriatric emergency department visits. Ulus Travma Acil Cerrahi Derg. 2024;30(4):271-5.
45. Singh S. Ottawa Ankle Rule: An Indian Perspective. J Foot Ankle Surg (Asia Pacific). 2015;2(1):8-12.
46. Springer BA, Arciero RA, Tenuta JJ, Taylor DC. A prospective study of modified Ottawa ankle rules in a military population. Interobserver agreement between physical therapists and orthopaedic surgeons. Am J Sports Med. 2000;28(6):864-8.
47. Stiell IG, McKnight RD, Greenberg GH, McDowell I, Nair RC, Wells GA, et al. Implementation of the Ottawa ankle rules. Jama. 1994;271(11):827-32.
48. Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell I, Reardon M, et al. Decision rules for the use of radiography in acute ankle injuries. Refinement and prospective validation. Jama. 1993;269(9):1127-32.
49. Thakur AK, Kandel PR. Reliability of 'Ottawa Ankle Rules' in Acute Ankle and Midfoot Injuries. J Univ Coll Med Sci. 2022;10(01):33-6.
50. Verma S, Hamilton K, Hawkins HH, Kothari R, Singal B, Buncher R, et al. Clinical application of the Ottawa ankle rules for the use of radiography in acute ankle injuries: an independent site assessment. AJR Am J Roentgenol. 1997;169(3):825-7.
51. Wang X, Chang SM, Yu GR, Rao ZT. Clinical value of the Ottawa ankle rules for diagnosis of fractures in acute ankle injuries. PLoS One. 2013;8(4):e63228.
52. Wynn-Thomas S, Love T, McLeod D, Vernall S, Kljakovic M, Dowell A, et al. The Ottawa ankle rules for the use of diagnostic X-ray in after hours medical centres in New Zealand. N Z Med J. 2002;115(1162):U184.

Table 1: Characteristics of the included studies

Study	Country	Age* (year)	Gender M%	Gold Standard	Sensitivity	Specificity
Ahmad et al. (12); 2025	Pakistan	36.73 ± 6.7	67.13	Radiography	0.56	0.95
Ak et al. (13); 2022	Turkey	33.1 ± 16.3	53.60	Radiography	0.97	0.48
Anyanwu et al. (14); 2025	Nigeria	37.4 ± 11	54.40	Radiography	1	0.37
Auley et al. (15); 1998	France	34 [18–90]	8	Radiography	0.98	0.45
Avinca et al. (16); 2023	Turkey	30.93 ± 13.154	36	Radiography	0.85	0.82
Beceren et al. (17); 2013	Turkey	30.3 ± 13.2	60.40	Radiography	0.74	0.68
Broomhead et al. (18); 2003	Australia	34.7 [18.1–84.8]	51.70	Radiography	1	0.15
Can et al. (19); 2008	Switzerland	Fracture: 51 ± 21; No fracture: 38 ± 17	Fracture: 39; No fracture: 56	Radiography	1	0.21
Daş et al. (20); 2016	Turkey	Mean: 37.46 [18–85]	61.20	Radiography	0.98	0.44
Derksen et al. (a) (21); 2005	Netherlands	Above 18	NR	Radiography	0.93	0.39
Derksen et al. (b) (21); 2015	Netherlands	Mean: 37 [18–73]	46.80	Radiography	0.97	0.29
Dwivedi et al. (22); 2014	Nepal	35.78 ± 11.39	54.90	Radiography	1	0.036
Fiesseler et al. (23); 2004	USA	37 ± 16	33	Radiography	0.92	0.47
Glas et al. (24); 2002	Netherlands	35 ± 14	50	Radiography	89	26
Gomes et al. (9); 2020	Australia	38 ± 18.3	50.40	Radiography	0.59	0.37
Jia et al. (25); 2025	China	Above 18	NR	Radiography	0.98	0.26
Kelly et al. (26); 1994	New Zealand	NR	NR	Radiography	0.93	0.11
Keogh et al. (27); 1998	UK	NR	NR	Radiography	1	0.48
Leisey et al. (28); 2004	Saudi Arabia	Mean: 27 [19–46]	82	Radiography	1	0.4
Lucchesi et al. (29); 1995	USA	Mean: 38 [18–91]	44.60	Radiography	0.94	0.15
Makiev et al. (30); 2025	Greece	39.6 ± 15.8	43	Radiography	0.94	0.15
Marinelli et al. (31); 2007	Italy	31.8 ± 15.9	64.90	Radiography	1	0.43
McBride et al. (32); 1997	Canada	No fracture: 30.90 ± 12.27; Fracture: 41.21 ± 17.87	No fracture: 54.6; Fracture: 38.2	Radiography	0.97	0.3
Morais et al. (33); 2021	Portugal	46.1 ± 18.7	34	Radiography	1	0.26
Murphy et al. (Geriatric) (34); 2020	USA	76.2 ± 8.1	20.20	Radiography	0.97	0.33
Murphy et al. (non-geriatric) (34); 2020	USA	39.0 ± 13.3	46.30	Radiography	0.98	0.6
Özbay et al. (35); 2021	Turkey	57.2 ± 20.9	51.60	Radiography	0.8	0.63
Papacostas et al. (36); 2001	Greece	29 ± 9.62	NR	Radiography	1	0.3
Perry et al. (37); 1999	UK	Above 18	NR	Radiography	0.93	0.46
Pigman et al. (38); 1994	USA	Fracture: 38; Non-fracture: 34	Fracture: 43; Non-fracture: 50	Radiography	1	0.19
Pijnenburg et al. (39); 2002	Netherlands	35 ± 14	50	Radiography	0.97	0.38
Pires et al. (40); 2014	Brazil	Mean: 37.4 [18–88]	40.90	Radiography	0.97	0.07
Rosin et al. (41); 1999	South Korea	Mean: 24.9 [19–41]	80.50	Radiography	0.7	0.73
Salt et al. (42); 1997	UK	Above 18	NR	Radiography	1	0.31
Santelli et al. (43); 2008	Italy	31.8 ± 15.9	NR	Radiography	1	0.46
Seyhan et al. (44); 2024	Turkey	73.6 ± 6.9	26.20	Radiography	0.98	0.86
Singh et al. (45); 2015	India	Mean: 34.76 [18–68]	61.30	Radiography	0.95	0.97
Springer et al. (46); 2000	USA	Mean: 20.25 [18–25]	80.39	Radiography	1	0.46
Stiell et al. (a) (47); 1994	Canada	36 ± 16	47	Radiography	1	0.5
Stiell et al. (b) (48); 1993	Canada	35 ± 15	52	Radiography	1	0.39
Thakur et al. (49); 2022	Nepal	Above 18	NR	Radiography	1	0.47
Verma et al. (Prospective) (50); 1997	USA	Above 18	NR	Radiography	0.98	0.2
Verma et al. (Retrospective) (50); 1997	USA	Above 18	NR	Radiography	1	0.02
Wang et al. (51); 2013	China	Mean: 36.6 [18–70]	44.26	Radiography/3D-CT	0.96	0.45
Wynn-Thomas et al. (52); 2002	New Zealand	Above 18	47	Radiography	1	0.47

*: Data are presented as mean ± standard deviation, median (interquartile range), or mean (range).
M: Male; NR: Not reported; 3D-CT: 3-dimensional computed tomography.

Table 2: Quality appraisal of the included studies

Study	Risk of bias				Applicability concerns		
	Patient selection	**Index test**	**Reference standard**	**Flow and timing**	**Patient selection**	**Index test**	**Reference standard**
Ahmad et al.	☹	?	?	?	☺	☺	☺
Ak et al.	☹	☺	☺	☺	☹	☺	☺
Anyanwu et al.	☹	?	☺	☺	☺	☺	☺
Auley et al.	☹	☺	☺	☺	☺	☺	☺
Avinca et al.	☹	☹	☹	?	?	☺	☺
Beceren et al.	☹	☺	☺	☺	☺	☺	☺
Broomhead et al.	☹	☺	☺	☺	☺	☺	☺
Can et al.	☹	☺	☺	☺	☹	☺	☺
Daş et al.	☹	☺	☺	☺	☹	☺	☺
Derksen et al. (a)	☹	☺	?	☺	☺	☺	☺
Derksen et al. (b)	☹	☺	☺	☺	☺	☺	☺
Dwivedi et al.	☹	?	?	☺	☹	☺	☺
Fiesseler et al.	☹	☺	☺	☹	☹	☺	☺
Glas et al.	☹	☺	☺	☺	☺	☺	☺
Gomes et al.	☹	☹	?	?	☹	☹	☺
Jia et al.	☺	☺	☺	☺	☺	☺	☺
Kelly et al.	?	?	?	?	☺	☺	☺
Keogh et al.	☺	?	?	?	☺	☺	☺
Leisey et al.	☺	☺	☺	?	☺	☺	☺
Lucchesi et al.	☺	☺	☺	☺	☺	☺	☺
Makiev et al.	☺	?	☺	☺	☺	☺	☺
Marinelli et al.	☺	☺	☺	☺	☺	☺	☺
McBride et al.	☹	?	?	☹	☺	☺	☺
Morais et al.	☺	☺	☺	☺	☺	☺	☺
Murphy et al. (Geriatric)	☹	?	?	☹	☺	☺	☺
Murphy et al. (Non-G)	☹	?	?	☹	☺	☺	☺
Özbay et al.	☺	☺	☺	☺	☺	☺	☺
Papacostas et al.	☺	?	☺	☺	☺	☺	☺
Perry et al.	☺	☺	☺	☺	☺	☺	☺
Pigman et al.	☺	☺	☺	☺	☺	☺	☺
Pijnenburg et al.	☺	☺	☺	☺	☺	☺	☺
Pires et al.	☺	?	☺	☺	☺	☺	☺
Rosin et al.	☹	?	?	☹	☺	☺	☺
Salt et al.	☺	☹	?	☹	☺	☺	☺
Santelli et al.	☺	?	?	☺	☺	☺	☺
Seyhan et al.	☺	☺	?	☺	☺	☺	☺
Singh et al.	☺	☺	?	?	☺	☺	☺
Springer et al.	☺	☺	☺	☺	☺	☺	☺
Stiell et al. (a)	☺	☺	☺	☺	☺	☺	☺
Stiell et al. (b)	☺	☺	☺	☺	☺	☺	☺
Thakur et al.	☺	☺	☺	☺	☺	☺	☺
Verma et al. (Pros)	☹	?	?	☹	☺	☺	☺
Verma et al. (Retro)	☹	?	?	☹	☺	☺	☺
Wang et al.	☺	☺	?	?	☺	☺	☺
Wynn-Thomas et al.	☺	☺	?	?	☺	☺	☺

☺Low Risk; ☹High Risk; ? Unclear Risk. Pros: Prospective; Retro: Retrospective; Non-G: non geriatric.

Table 3: GRADE summary of certainty

<table>
<tr><th rowspan="2">Outcome</th><th rowspan="2">N</th><th rowspan="2">Design</th><th colspan="5">Factors that may decrease certainty of evidence</th><th colspan="3">Effect per 100 patients tested</th><th rowspan="2">Test accuracy</th></tr>
<tr><th>Risk of bias</th><th>Indirectness</th><th>Inconsistency</th><th>Imprecision</th><th>Publication bias</th><th>PTP 10%</th><th>PTP 15%</th><th>PTP 20%</th></tr>
<tr><td>True positives (patients with Ankle Fractures)</td><td rowspan="2">45 studies 3066 patients</td><td rowspan="2">Cross-sectional (cohort type accuracy study)</td><td rowspan="2">Serious</td><td rowspan="2">Not serious</td><td rowspan="2">Serious</td><td rowspan="2">Not serious</td><td rowspan="2">All plausible residual confounding would reduce the demonstrated effect</td><td>9 (9 to 9)</td><td>14 (14 to 14)</td><td>18 (18 to 19)</td><td rowspan="2">⊕ ⊕ ⊕◯ Moderate</td></tr>
<tr><td>False negatives (patients incorrectly classified as not having Ankle Fractures)</td><td>1 (1 to 1)</td><td>1 (1 to 1)</td><td>2 (1 to 2)</td></tr>
<tr><td>True negatives (patients without Ankle Fractures)</td><td rowspan="2">45 studies 13437 patients</td><td rowspan="2">Cross-sectional (cohort type accuracy study)</td><td rowspan="2">Serious</td><td rowspan="2">Not serious</td><td rowspan="2">Serious</td><td rowspan="2">Not serious</td><td rowspan="2">All plausible residual confounding would reduce the demonstrated effect</td><td>32 (31 to 32)</td><td>30 (29 to 31)</td><td>28 (27 to 29)</td><td rowspan="2">⊕ ⊕ ⊕◯ Moderate</td></tr>
<tr><td>False positives (patients incorrectly classified as having Ankle Fractures)</td><td>58 (58 to 59)</td><td>55 (54 to 56)</td><td>52 (51 to 53)</td></tr>
</table>

N: number; PTP: pre-test probability; GRADE: Grading of Recommendations, Assessment, Development and Evaluation.

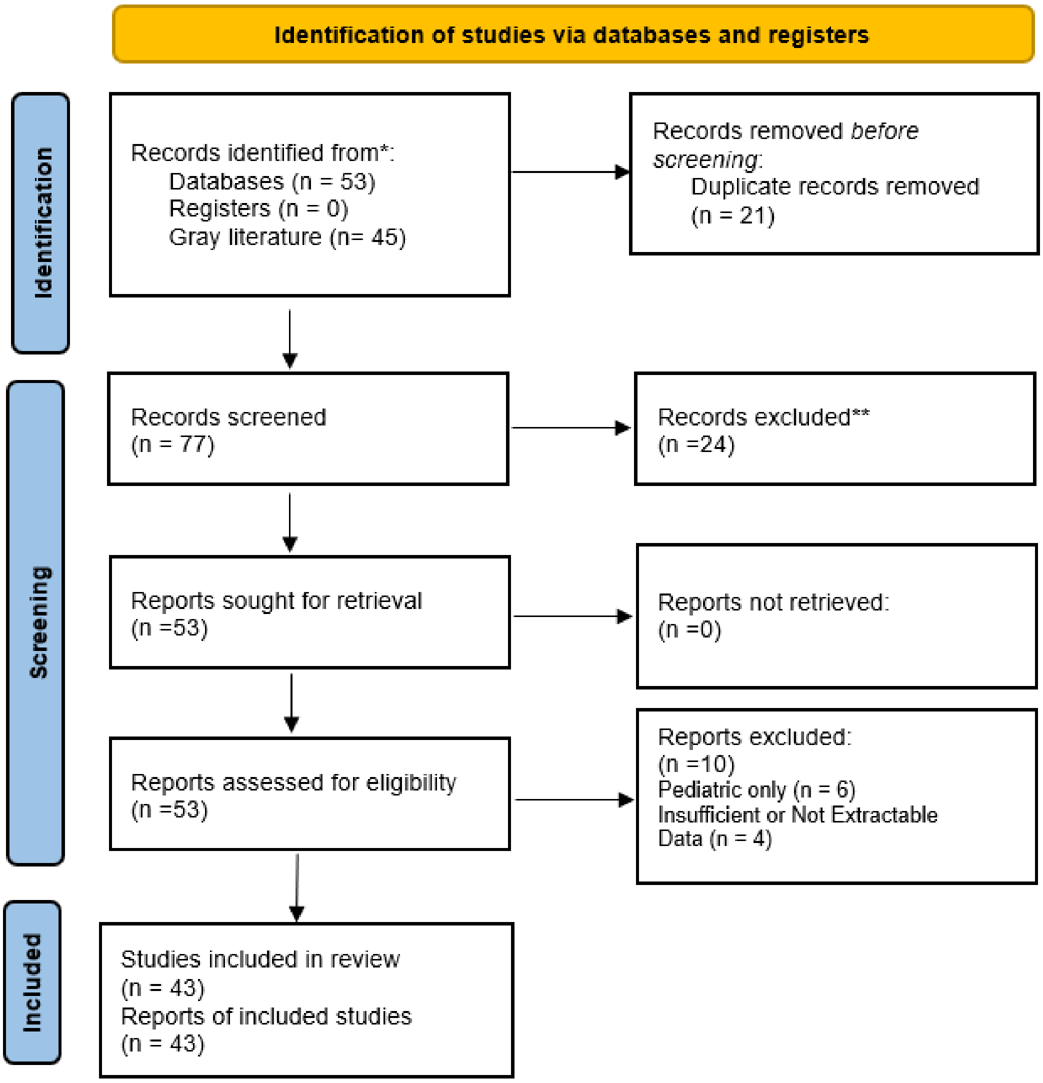

Figure 1: Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) flowchart of the included studies.

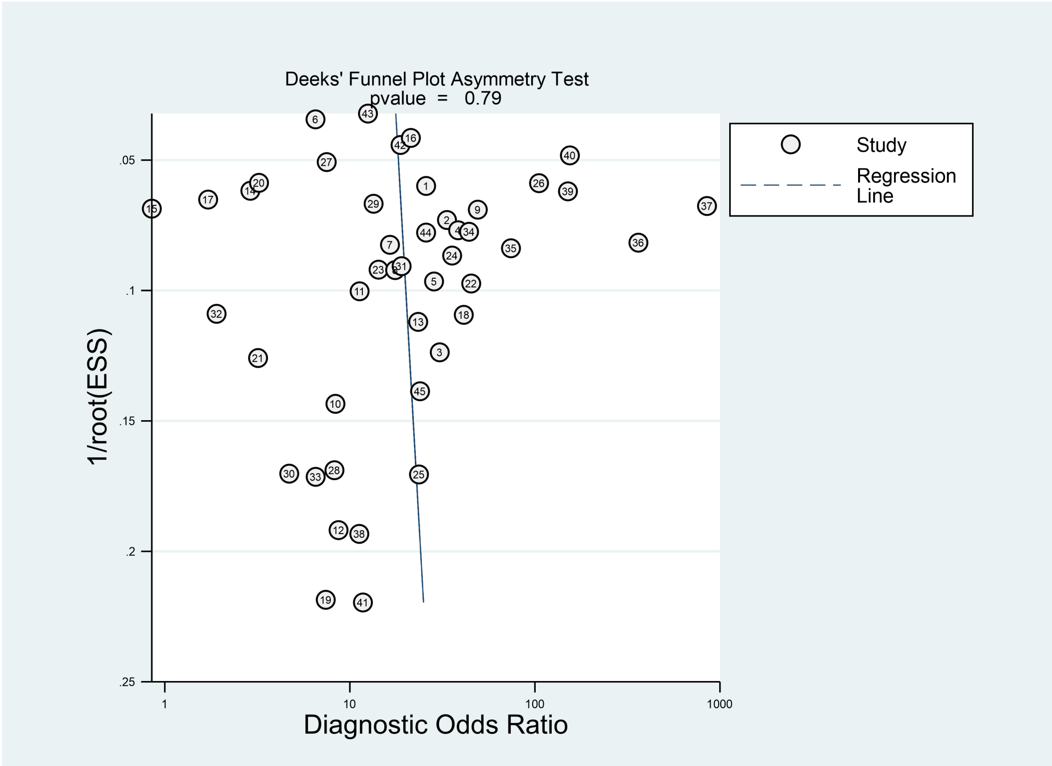

Figure 2: Deeks' funnel plot and test.

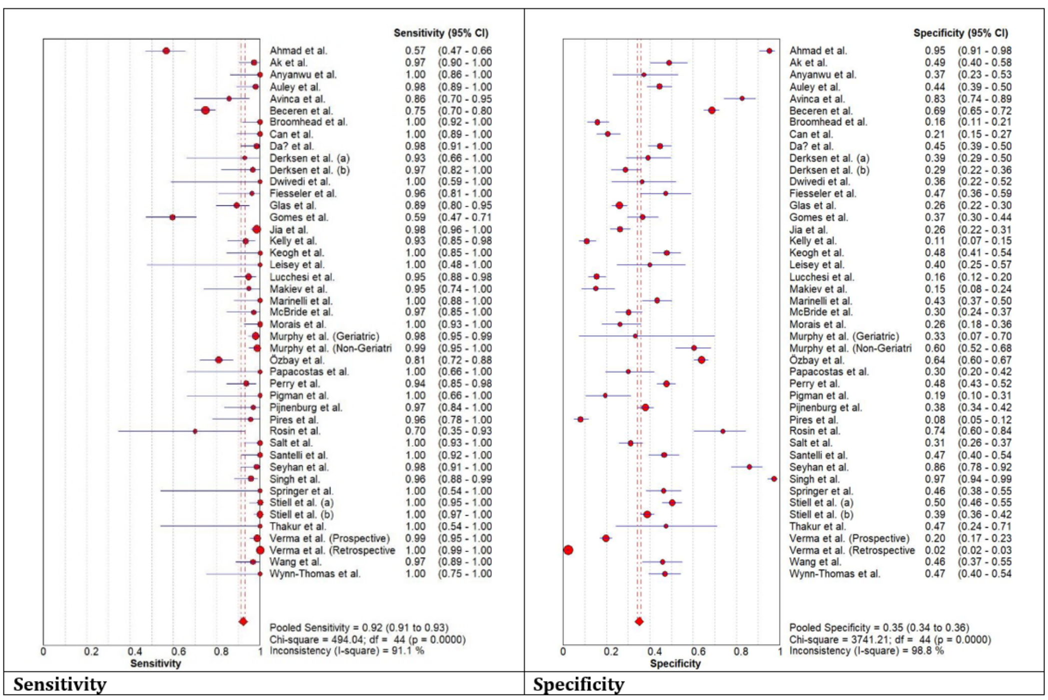

Figure 3: Pooled sensitivity and specificity of Ottawa Ankle Rules for detection of ankle fractures. CI: confidence interval.

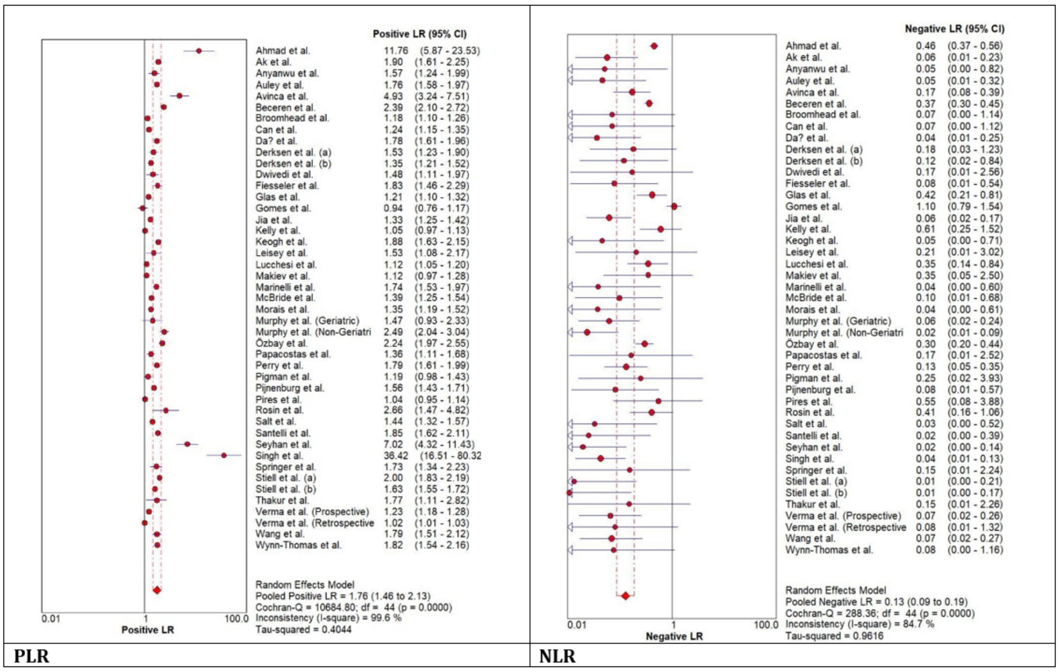

Figure 4: Pooled positive likelihood ratio (PLR) and negative likelihood ratio (NLR) of Ottawa Ankle Rules for detection of ankle fractures. CI: confidence interval; LR: likelihood ratio.

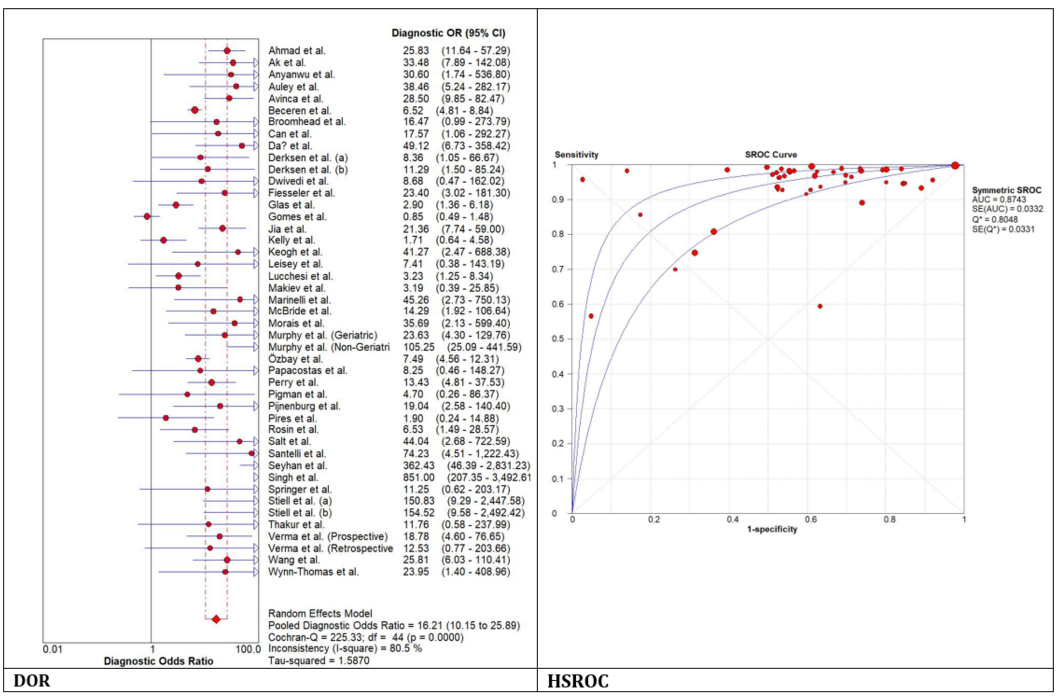

Figure 5: Pooled diagnostic odds ratio (DOR) and the area under the hierarchical summary receiver operating characteristic (HSROC) curve (AUC) of Ottawa Ankle Rules for detection of ankle fractures. CI: confidence interval; OR: odds ratio.

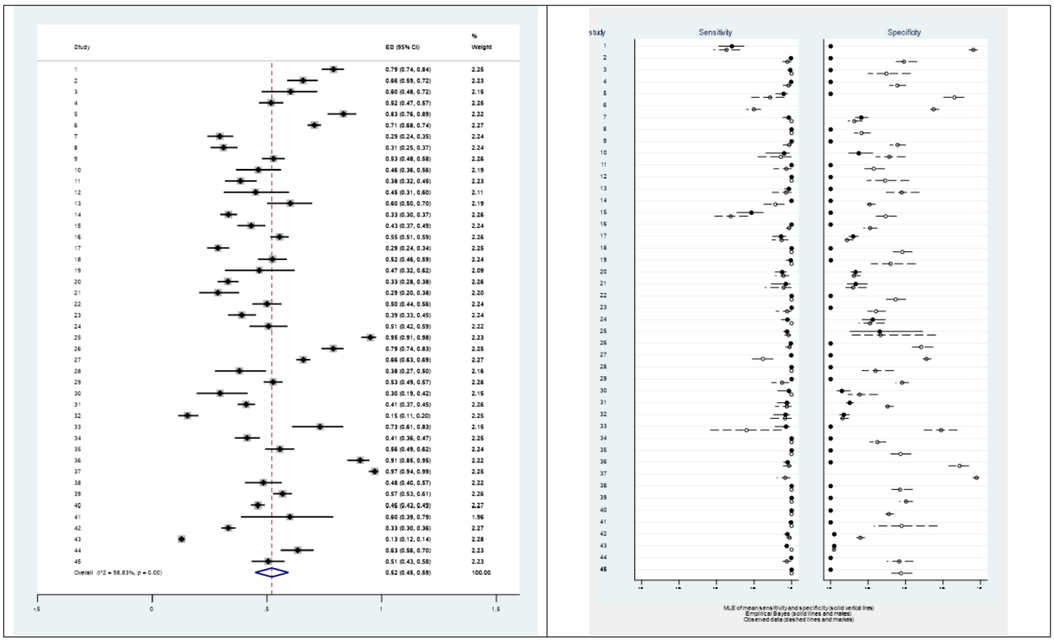

Figure 6: Pooled accuracy of Ottawa Ankle Rules for detection of ankle fractures (left); Forest plot of observed and empirical Bayes estimates of sensitivity and specificity across included studies (right). CI: confidence interval

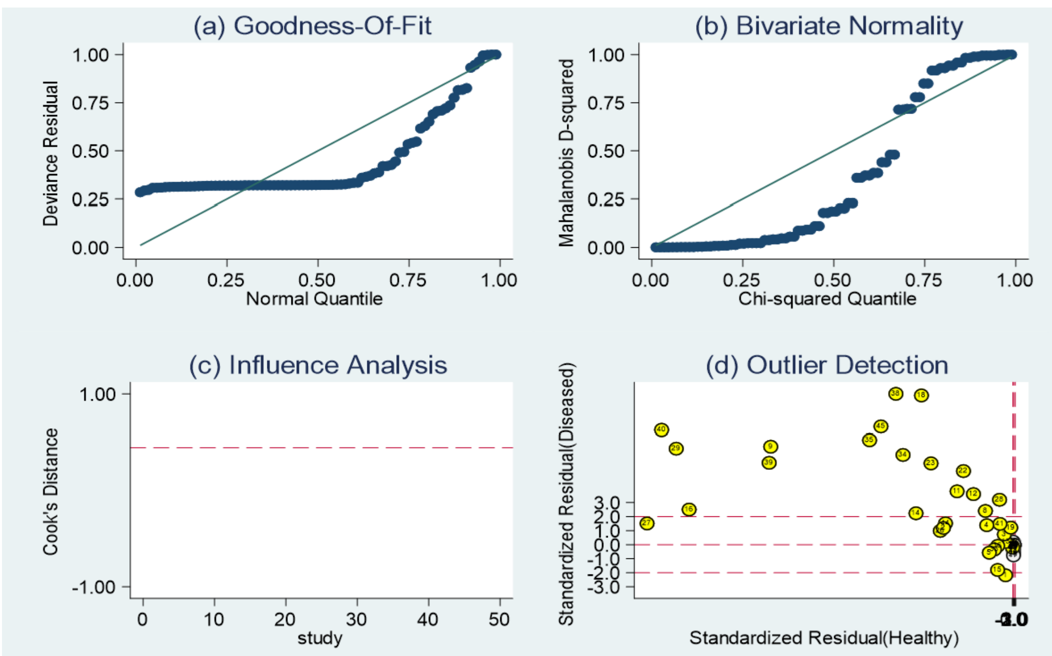

Figure 7: Goodness-of-fit and bivariate normality plots for model evaluation, along with influence analysis and outlier detection plots for the identification of influential and outlier studies.

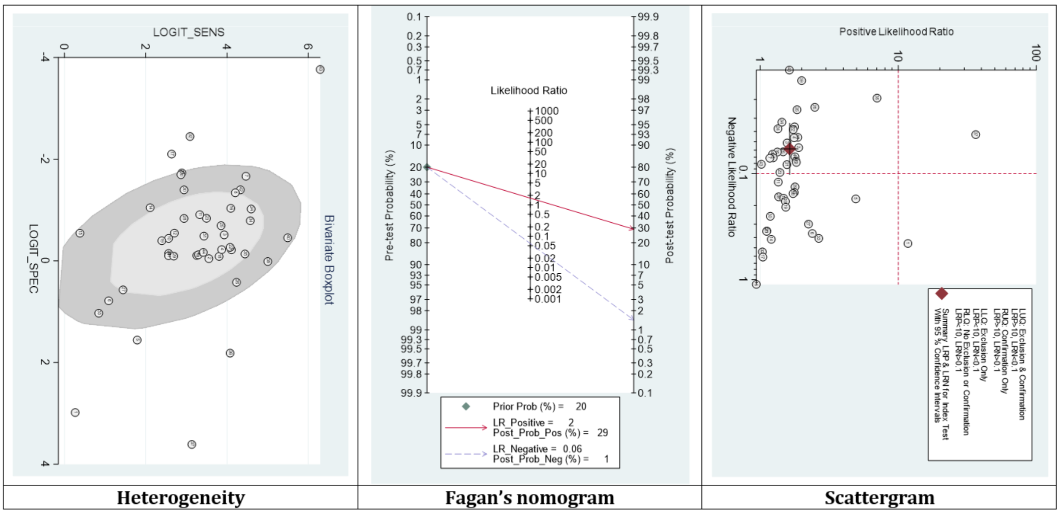

Figure 8: Boxplot for evaluation of heterogeneity and identifying outliers (left); Fagan's nomogram illustrating the effect of the Ottawa Ankle Rules on the post-test probability of ankle fractures (middle); Likelihood ratio (LR) scattergram illustrating the role of the Ottawa Ankle Rules in ruling out and confirming ankle fractures.

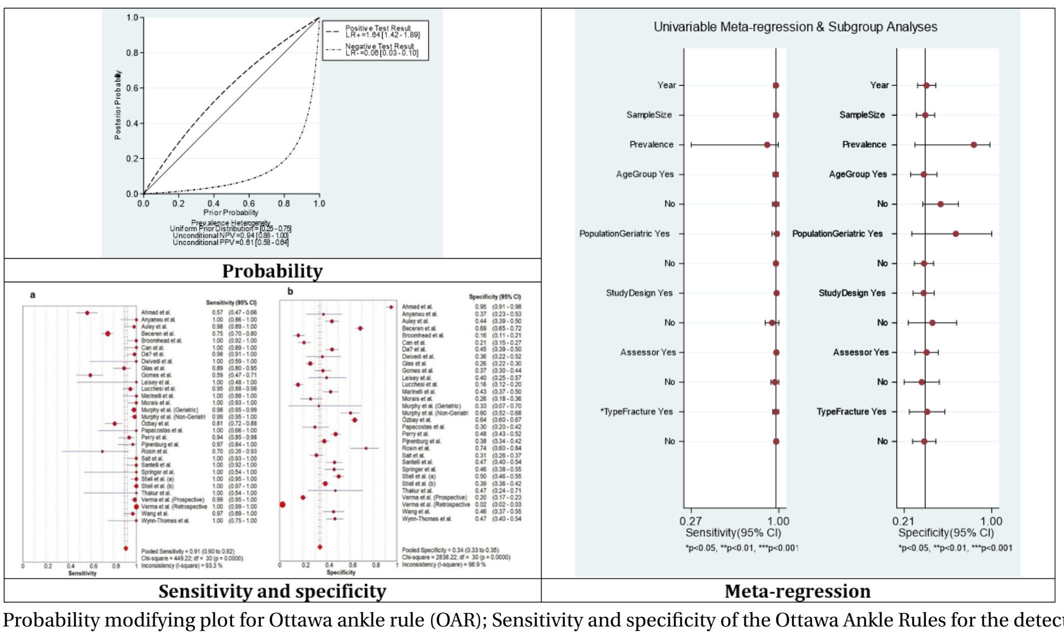

Figure 9: Probability modifying plot for Ottawa ankle rule (OAR); Sensitivity and specificity of the Ottawa Ankle Rules for the detection of ankle fractures based on studies that exclusively reported ankle fracture data or provided sufficient data for separate extraction of ankle fracture outcomes; Meta-regression based on the clinical and demographic covariates. Age Group Yes: Age≥35, Age Group No: Age<35; Population Geriatric Yes: Geriatric Population; Study Design Yes: Prospective Design, Study Design No: Retrospective Design; Assessor Yes: Physician, Assessor No: Other Assessors or mixed of different assessors; Type Fracture Yes: Ankle and Foot Fractures, Tye Fracture No: Only Ankle Fractures; CI: confidence interval; LR likelihood ratio.